A Comparative Study of Users' Microblogging Behavior on Sina Weibo and Twitter

Qi Gao[1], Fabian Abel[1], Geert-Jan Houben[1], Yong Yu[2]

[1] Web Information Systems, Delft University of Technology
[2] APEX Data & Knowledge Management Lab, Shanghai Jiaotong University
{q.gao,f.abel,g.j.p.m.houben}@tudelft.nl, yyu@apex.stju.edu.cn

Abstract. In this article, we analyze and compare user behavior on two different microblogging platforms: (1) Sina Weibo which is the most popular microblogging service in China and (2) Twitter. Such a comparison has not been done before at this scale and is therefore essential for understanding user behavior on microblogging services. In our study, we analyze more than 40 million microblogging activities and investigate microblogging behavior from different angles. We (i) analyze how people access microblogs and (ii) compare the writing style of Sina Weibo and Twitter users by analyzing textual features of microposts. Based on semantics and sentiments that our user modeling framework extracts from English and Chinese posts, we study and compare (iii) the topics and (iv) sentiment polarities of posts on Sina Weibo and Twitter. Furthermore, (v) we investigate the temporal dynamics of the microblogging behavior such as the drift of user interests over time.

Our results reveal significant differences in the microblogging behavior on Sina Weibo and Twitter and deliver valuable insights for multilingual and culture-aware user modeling based on microblogging data. We also explore the correlation between some of these differences and cultural models from social science research.

Key words: user modeling, microblogging, comparative usage analysis

1 Introduction

Microblogging services such as Twitter allow people to publish, share and discuss short messages on the Web. Nowadays, Twitter users publish more than 200 million posts, so-called *tweets*, per day[3]. In China, Sina Weibo[4] is leading the microblogging market since Twitter is unavailable. Both Sina Weibo and Twitter basically feature the same functionality. For example, both services limit the lengths of microposts to 140 characters and allow users to organize themselves in a follower-followee network, where people follow the message updates of other users (unidirectional relationship). Sina Weibo and Twitter provide (real-time) access to the microposts via APIs and therefore allow for investigating and analyzing interesting applications and functionality such as event detection [1, 2] or recommending Web sites [3].

By analyzing individual microblogging activities, it is possible to learn about the characteristics, preferences and concerns of users. In previous work, we therefore introduced a semantic user modeling framework for inferring user interests

[3] http://blog.twitter.com/2011/06/200-million-tweets-per-day.html
[4] http://www.weibo.com/

from Twitter activities and proved its efficiency in a news recommendation system [4]. In this paper, we extend this Twitter-based user modeling framework to also allow for sentiment analysis and user modeling based on Chinese microblog posts. We conduct, to the best of our knowledge, the first comparative study of the microblogging behavior on Sina Weibo and Twitter and relate our findings to theories and models from social science. The main contributions of our work can be summarized as follows.

- We extend our framework for user modeling based on usage data from microblogging services with functionality for sentiment analysis and semantic enrichment of Chinese microblog posts.
- We conduct intensive analyses based on more than 40 million microblog posts and compare the microblogging behavior on Sina Weibo and Twitter regarding five dimensions: (i) access behavior, (ii) syntactic content analysis, (iii) semantic content analysis, (iv) sentiment analysis, (v) temporal behavior.
- We relate our findings to theories about cultural stereotypes developed in social sciences and therefore explain how our insights can allow for culture-aware user modeling based on microblogging streams.

2 Related Work

Various types of research efforts have been conducted on Twitter data recently ranging from information propagation [5, 6] to applications such as Twitter-based early warning systems [1]. Furthermore, user modeling and personalization research started to study Twitter. Chen et al. investigate recommender systems on Twitter that consider social network features or the popularity of items in the Twitter network [3]. In previous work, we developed a Twitter-based user modeling framework for inferring user interests [4] and studied different applications that exploit the framework for personalization [7].

Research on cultural characteristics of user behavior on the Social Web has also been initiated. For example, Mandl [8] investigates how blog pages, especially the communication patterns between bloggers and commentators, from China differ from the ones from Germany. He correlates his findings to cultural dimensions proposed by Hofstede et al. [9]. Chen et al. analyze the tagging behavior of two user groups from two popular social music sites in China and Europe respectively [10] and observe differences between the two cultural groups, e.g. Chinese users have a smaller tendency to apply subjective tags but prefer the usage of factual tags. So far, there exists little knowledge about the differences and commonalities regarding the microblogging behavior of users from different cultural groups. Yu et al. compare popular trending topics on Sina Weibo with those on Twitter [11], but only compare global trends and do not study individual user behavior. In this paper, we close this gap: based on our extended user modeling framework, we conduct a large-scale analysis and comparison of users' microblogging behavior on Sina Weibo and Twitter.

3 Research Methodology and Evaluation Platform

In this section, we detail our research questions and present our enhanced user modeling environment that allows us to investigate the research questions.

3.1 Research Questions

Our research goal is to analyze and compare user behavior on Sina Weibo and Twitter to gain insights for user modeling on microblogging streams. Therefore, we investigate (1) how people access microblogging services, (2) the content, (3) semantics and (4) sentiment of microblog posts and (5) the temporal behavior of users' microblogging activities.

Analysis of Access Behavior Microblogging services such as Sina Weibo and Twitter can be accessed via different client applications from both mobile devices and desktop devices. User behavior that can be observed on a microblogging service may be influenced by the way in which a user accesses the service. We thus first study the following research questions:

- *RQ1:* How do people access Sina Weibo and Twitter respectively to publish microposts?
- *RQ2:* To what extent do individual users access a microblogging service from different client applications?

Syntactic Content Analysis Both Sina Weibo and Twitter limit the length of posts to 140 characters. This limitation impacts the writing style of microblog users and may result in characteristic usage patterns that we would like to compare between Sina Weibo (Chinese) and Twitter (English):

- *RQ3:* How does the usage of hashtags, URLs and other syntactic patterns (e.g. punctuation) differ between Sina Weibo and Twitter for both (i) the entire user population and (ii) individual users?
- *RQ4:* To what extent is the usage of hashtags and URLs influenced by the users' access behavior?

Semantic Content Analysis To better understand the meaning of the messages that users post on microblogging services, we analyze the semantics and investigate the following aspects:

- *RQ5:* What kind of topics and concepts do users mention and discuss on Sina Weibo and Twitter respectively?
- *RQ6:* To what extent do the types of concepts that users mention in their posts depend on the client applications via which they publish their posts?

Sentiment Analysis Microblogs allow users to express and discuss their opinions about topics that people are concerned with. We therefore analyze the sentiment of Chinese and English messages and study the following questions:

- *RQ7:* To what extent do users reveal their sentiment on Sina Weibo and Twitter respectively?
- *RQ8:* To what extent does the sentiment correlate with the type of topics and concepts that people mention in their Sina Weibo and Twitter messages?

Analysis of Temporal Behavior The users' microblogging behavior may change over time and may, for example, differ between working hours and leisure time. Therefore, we investigate the following research questions:

- *RQ9:* How does the posting behavior of users, particularly regarding the type of topics that the users mention, change between weekdays and weekends on Sina Weibo and Twitter?
- *RQ10:* How do individual user interests change over time in the two microblogging services?

3.2 Evaluation Platform

Extended User Modeling Framework for Microblogging Services. In previous work, we developed a Twitter-based user modeling framework for inferring user interest from tweets [4, 7]. Our framework monitors Twitter activities of a user and enriches the semantics of her Twitter messages by extracting meaningful concepts and topics (e.g. named entities) from the messages' content and by linking posts to external relevant Web resources such as new articles. Different weighting schemes such as time-sensitive or term-frequency-based functions allow for estimating to what extent a user might be interested in a given concept at a particular point in time. The generated user profiles can therefore be considered as a set of weighted semantic concepts.

In this paper, we extend our framework with three core features: (1) functionality for monitoring microblogging activities and collecting microposts published on Sina Weibo, (2) named entity recognition for Chinese microposts and (3) sentiment analysis for both Chinese and English microposts. We use ICTCALS[5] as part-of-speech tagger for Chinese text and extract named entities such as locations, organizations and persons from Chinese posts. We implemented a baseline approach to analyze the sentiment of Chinese and English microposts as proposed in [12]. Given these additional features, we are able to apply the same user modeling techniques on both microblogging services Sina Weibo and Twitter and can therefore analyze and compare user characteristics and behavior on the Asian and Western microblogging platforms.

Data Collection Given the framework, we collected microposts over a period of more than two months via the Sina Weibo Open API and the Twitter Streaming API respectively. For Twitter, we started from a seed set of 56 Twitter users and then we gradually extended this set in a snowball manner. Overall, we collected more than 24 million tweets published by more than 1 million users. For Sina Weibo, since it does not provide functionality similar to Twitter's Streaming API, we monitored the most recent public microposts and finally collected more than 22 million microposts published by more than 6 million users. Twitter posts and Sina Weibo posts were then processed by our framework in order to enrich the semantics of the posts (e.g. entity extraction, sentiment analysis). To better understand the behavior on the level of individual users, we extracted a sample of 1200 active Twitter users (who post in English) and 2616 active Sina Weibo users. The majority of the Twitter users (more than 80%) is – according to their Twitter profile – from the United States while the great majority of the Sina Weibo users (more than 95%) is located in China. For a detailed description on the dataset characteristics we refer the reader to [4] and [2] respectively.

4 Analysis of User Behavior on Sina Weibo and Twitter

Based on the more than 40 million posts that we collected from Sina Weibo and Twitter and processed with our user modeling framework, we study the users' behavior on the two platforms and answer the research questions regarding the five dimensions ranging from access behavior to temporal behavior.

type of access	fraction of posts Weibo	Twitter
posted on a Web or desktop application	54.9	66.2
posted on a mobile application	45.1	33.8
primary product of microblogging activity	90.6	96.7
byproduct of an activity on another platform	9.4	3.3

Table 1. Number of posts published via different categories of access clients

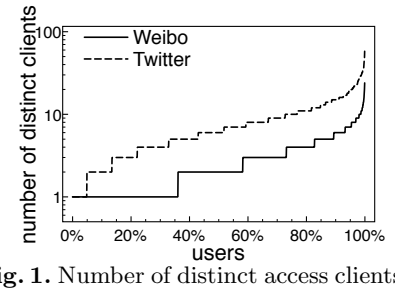


Fig. 1. Number of distinct access clients for individual users

4.1 Analysis of Access Behavior

Results We first analyzed the most popular client applications that people use to publish posts on Sina Weibo and Twitter. On both platforms, the Web interface is the most popular way to access the microblogging services: 43.1% of the posts are published via the Web on Sina Weibo and 38.5% on Twitter. Other popular clients on Sina Weibo are mainly designed for mobile devices such as the iPhone (7.6%) and Nokia devices (9.4%). Among the most popular Twitter clients are many desktop-based applications such as *TweetDeck*, via which 10.7% of the posts are published. Moreover, we observe on both platforms that people publish posts that are rather byproducts of activities the users perform on other platforms. For example, 1.3% of the posts in our Twitter dataset are published via *Twitterfeed*, an application that allows for publishing announcements on a user's Twitter timeline whenever she publishes a new blog article.

In Table 1, we overview the type of client applications that people use to publish microblog posts. We therefore manually categorized the 50 most popular clients, that generate more than 90% of the posts on both microblogging services. We observe that the fraction of posts that are published via mobile devices is significantly higher on Sina Weibo (45.1%) in comparison to Twitter (33.8%). Furthermore, we discover that the fraction of posts which are rather byproducts of other Web activities of the users – hence where the intent of the actual user activity was not targeted towards Sina Weibo or Twitter – is almost three times higher on Sina Weibo (9.4%) than on Twitter (3.3%).

In Fig. 1, we plot for each of the sample users the number of distinct applications which they utilize for publishing microposts. We see that on Twitter more than 95% of the people use more than one client application while on Sina Weibo around 65% of the users switch between different clients.

Findings From the results above, we conclude the analysis of access behavior with two main findings, referring to the research questions *RQ1* and *RQ2*:

– *F1:* On both platforms, the major way to accessing the microblogging services is via the official Web interfaces or desktop-based applications. Chinese users seem to differ from the English-spoken Twitter users regarding two core aspects: (i) they use mobile applications more extensively and (ii) publish microposts more often as a byproduct of their other Social Web activities.

[5] `http://ictclas.org/`

syntactic characteristics proportion of posts posts that contain:	Weibo	Twitter
hashtags	6.3%	20.0%
URLs	14.8%	29.1%
question marks "?"	9.9%	18.6%
exclamation marks "!"	26.1%	20.7%
"?" and "!"	3.1%	3.5%

Table 2. Comparison of syntactic content analysis

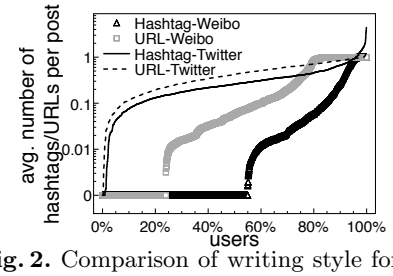


Fig. 2. Comparison of writing style for individual users

– *F2:* The results regarding the individual users' access behavior illustrate that Twitter users switch between different clients more often than the users on Sina Weibo. This difference in behavior could be explained by the lower overall number of valuable Sina Weibo client applications (e.g. in our dataset: 3015 different Sina Weibo clients versus 5468 Twitter clients).

4.2 Syntactic Content Analysis

Results In Table 2, we compare the syntax of messages posted on Sina Weibo and Twitter and particularly the usage of hashtags and URLs. Overall, 20% of the Twitter messages contain hashtags and 29.1% of the tweets feature a URL. Therefore, the usage of hashtags and URLs on Twitter is 3.2 times and 1.97 times respectively more intensive than on Sina Weibo. The analysis of special characters implies that users on Twitter ask more than twice as many questions than users on Sina Weibo (see question marks in Table 2). In contrast, Sina Weibo users make more extensive use of exclamation marks and therefore more often put extra emphasis on their statements.

To further analyze the usage of hashtags and URLs, we also plot for each individual user in our samples the average number of hashtags and URLs per post. From Fig. 2, we infer that a considerably high fraction of Sina Weibo users does not mention hashtags or URLs at all. For 55% of the Chinese microbloggers on Sina Weibo, we did not observe any hashtag. In contrast, on Twitter the people make more frequently use of hashtags or URLs. For example, for more than 85% of the Twitter users, the average number of hashtags per post is at least 0.1, i.e. at least every tenth micropost mentions a hashtag, and 3.9% of the users mention, on average, even more than one hashtag per tweet.

In Table 3 we analyze the influence of the access behavior (see Sect. 4.1) on the usage of hashtags and URLs. For both services, we observe that the usage of hashtags and URLs decreases slightly when people publish microposts from their mobile devices instead of their desktop computers. This difference is more significant on Sina Weibo. For example, on Sina Weibo the number of posts that contain a URL and are issued from a desktop application (17.8%) is more than three times higher than the one for mobile devices (5.2%). On Twitter, the usage of URLs on desktop devices is only 1.57 times higher than on mobile devices. Regarding the type of activity that a user performed to publish a micropost, we observe that 97.9% of the tweets that were generated as byproducts of other activities (e.g. publishing an article in a blog or "check-in" activities on *Foursquare*) contain URLs. In contrast, for the conventional microblogging, only 25.3% of the Twitter messages contain URLs. A similar increase can be

Table 3. Impact of the access behavior on the syntactic characteristics of microposts

Syntactic characteristics posts that contain:	proportion of posts Weibo		Twitter	
	Desktop/Mobile	*Microblog/Byproduct*	*Desktop/Mobile*	*Microblog/Byproduct*
hashtags	6.5%/3.5%	3.8%/17.9%	20.7%/18.6%	19.9%/21.3%
URLs	17.8%/5.2%	5.7%/73.5%	31.6%/20.1%	25.3%/97.9%

observed on Sina Weibo. The number of hashtags is slightly less influenced by the type of activity that caused a micropost (see Table 3).

Findings Given the results above, we can answer *RQ3* and *RQ4* as follows:

- *F3:* Overall, the results show that hashtags and URLs are less frequently applied on Sina Weibo than on Twitter. This finding holds for both (i) the entire user population and (ii) individual users. In fact, we observe that a large fraction of users on Sina Weibo does not make use of hashtags which implies that hashtag-based user profiles, as discussed in [4], or topic modeling based on hashtags, as proposed by Romero et al. [6] do not seem to be appropriate on Sina Weibo. The usage statistics regarding question marks indicate that Twitter users ask twice more questions than Sina Weibo users.
- *F4:* The usage of hashtags and URLs is moreover influenced by the access behavior. We discover that (i) users are more likely to use hashtags and URLs when they post messages via desktop applications than via mobile applications. Furthermore, (ii) whenever messages are published as a byproduct of another activity – where the primary intention of the user is rather the promotion of an activity that the user performed on another platform – the probability that a micropost contains a hashtag or URL increases. A large fraction of these *byproduct microposts* seems to be automatically generated based on the activity the user performed on another platform. For user modeling those posts offer means to further contextualize the microblogging activities by following the URLs that are contained in the posts (cf. [4]).

4.3 Semantic Content Analysis

Results Based on the semantic enrichment provided by our user modeling framework, we analyze and compare the types of concepts and topics that people mention in their microposts on Sina Weibo and Twitter respectively. In Table 4 we compare the usage of three types of entities (location, people and organization). Most of the extracted semantic concepts refer to locations (e.g. cities, points of interests): 58.4% for Sina Weibo and 44.6% for Twitter. On Twitter, posts that refer to organizations (e.g. companies, institutions) are more than four times more likely to appear than on Sina Weibo. Examples of entities that were trending on Twitter include different types of entities such as “Mubarak” (person), the former president of Egypt, or “Republican Party” (organization). In contrast, the most popular entities on Sina Weibo are related to locations such as “Beijing” or “United States”.

Fig. 3 depicts the average number of entities that can be extracted per post for the individual users in our sample. For 24.8% of the Sina Weibo users, one can detect, on average, more than one entity per post. Moreover, the fraction of users for whom no entity can be extracted is 7.9% in contrast to 10.1% on Twitter. The semantics of the users’ messages posted on Sina Weibo are therefore easier to deduce than on Twitter. Based on a comparison of a sample of

type of posts	proportion of posts Weibo	Twitter
Location	58.4%	44.6%
Organization	3.3%	16.0%
Person	38.3%	39.4%

Impact of the access behavior on the type of concepts mentioned in the posts

	Desktop/ Mobile	*Microblog/ Byproduct*	*Desktop/ Mobile*	*Microblog/ Byproduct*
Location	11.2%/6.6%	15.5%/4.0%	9.3%/8.4%	8.9%/13.7%
Organization	0.7%/0.6%	0.9%/0.4%	3.5%/2.9%	3.3%/4.5%
Person	12.4%/12.3%	17.4%/4.9%	8.1%/6.7%	7.6%/8.7%

Table 4. Semantic analysis overall and impact of access behavior on the semantics.

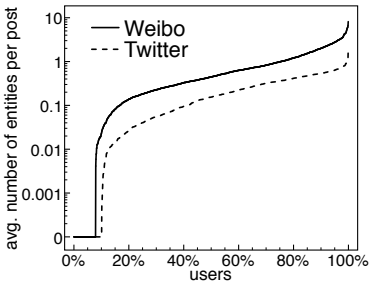


Fig. 3. Semantic analysis for individual users

individual Chinese and English microposts, we hypothesize that this is caused by the expressivity of the Chinese language: while Twitter users are often forced to leave out entities or use abbreviations to refer to entities, Sina Weibo users can exploit the 140 characters more effectively.

Table 4 illustrates how the access behavior influences the semantics of the microposts. When users publish posts from their mobile devices, then it becomes less likely, in comparison to access via desktop (tailored Web) applications, that a message mentions an entity. For microposts that are byproducts of other Web activities (e.g. activities on *Foursquare*), we observe that it becomes more likely that entities and particularly location entities are mentioned in a post on Twitter. In contrast, on Sina Weibo users mention more entities in context of their standard microblogging activities.

Findings The results of the analysis illustrate the commonalities and differences regarding the semantic meaning of the microposts that users publish on Sina Weibo and Twitter respectively (see *RQ5* and *RQ6* in Sec. 3.1):

- *F5:* The topics that users discuss on Sina Weibo are to a large extent related to locations and persons. In contrast to Twitter, users on Sina Weibo avoid talking about organizations such as political parties or other institutions. Overall, the semantics of Sina Weibo messages can be better extracted than the semantics of tweets. Consequently, when modeling the microblogging activities for individual users, entity-based user profiles [4] can more successfully be generated for Sina Weibo users: for 92.1% of them one can identify at least one entity of interest in comparison to 89.9% on Twitter.
- *F6:* The type of applications via which users access the microblogging services, affects the occurrence of semantic concepts in the microposts. On mobile devices people tend to mention less entities than on desktop devices. Furthermore, microposts on Twitter are more likely to mention entities and locations particularly if the post was generated as a byproduct of an activity performed on another platform.

4.4 Sentiment Analysis

Results The sentiment analysis provided by our framework classifies microblog posts as either positive, negative or neutral. Overall, 83.4% and 82.4% of the Sina Weibo and Twitter posts respectively were classified as neutral. Table 5 overviews the sentiment polarities of those posts that have been classified as

type of posts	proportion of positive/negative posts Weibo	Twitter
Overall posts	78.8%/21.2%	70.5%/29.5%
posts that mention certain types of entities:		
Location	82.7%/17.3%	65.6%/34.4%
Organization	78.5%/21.5%	70.1%/29.9%
Person	82.8%/17.2%	65.7%/34.3%

Table 5. Sentiment expressed in (i) overall posts and (ii) posts that mention certain types of topics

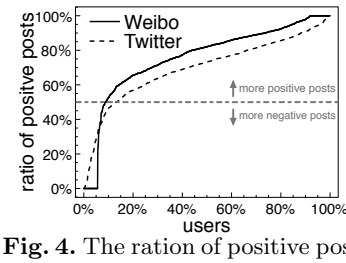


Fig. 4. The ration of positive posts on two microblogging services

positive or negative. On Sina Weibo the portion of positive posts (78.8%) is clearly higher than on Twitter (70.5%). In Fig. 4 we plot the ratio of positive posts with respect to all posts, which either have a positive or negative sentiment, for individual users: 92.5% of the users publish more positive messages than negative ones on Sina Weibo in comparison to 86.4% for the Twitter users. On Sina Weibo, we also discover a considerable fraction of users for whom the non-neutral posts are always positive (8.0%) or always negative (5.6%).

In Table 5 we moreover analyze the sentiment revealed in the microposts that mention certain types of entities. Again, the proportion of positive posts exceeds the proportion of negative posts clearly and Sina Weibo users tend to be more positive towards mentioned entities than Twitter users. Interestingly, whenever locations or persons are mentioned in Sina Weibo messages then the likelihood that the post is positive increases on Sina Weibo (from 78.8% to 82.7% and 82.8% respectively) while on Twitter the opposite can be observed (decrease from 70.5% to 65.6% and 65.7% respectively).

Findings Regarding the research questions *RQ7* and *RQ8* about the sentiment that users express in their microposts, we conclude the following:

- *F7:* We observe that on both platforms there are significantly more positive posts than negative ones. Moreover, users on Sina Weibo have a stronger tendency to publish positive messages than Twitter users. In fact, the probability for positive messages is 11.8% higher on Sina Weibo than on Twitter.
- *F8:* The sentiment that is expressed in microposts correlates with the type of concepts that are mentioned in the posts. On Sina Weibo posts that mention locations or persons are more likely to be positive than posts containing organizations. While on Twitter, the opposite can be observed: people talk more positively about organizations than about persons or locations.

4.5 Analysis of Temporal Behavior

Results In Table 6 we first compare the posting behavior of users between working days and weekend days by calculating the ratio between the average number of posts per day published during the weekends (Saturday-Sunday) and the one during the week (Monday-Friday). For Sina Weibo this ratio is 1.19, which means that Sina Weibo user publish, on average, 19% more messages per day on the weekend than they do during the week. On the other hand, the users on Twitter publish, on average, 11% less posts during the weekend. Therefore, it seems that microblogging in China has not penetrated the daily (possibly work-related) routines as strongly as it does in Western countries.

posts per weekend day / posts per weekday	Weibo	Twitter
Overall posts	1.19	0.89
posts that mention certain types of entities:		
Location	0.81	1.05
Organization	1.50	0.91
Person	1.19	0.97

Table 6. Ratio between weekend posts and weekday posts = the average number of posts per day on a weekend divided by the average number of posts per weekday

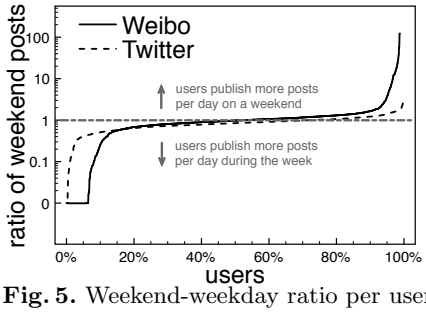


Fig. 5. Weekend-weekday ratio per user

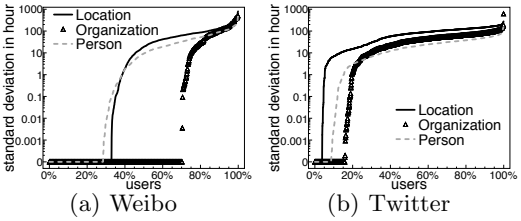
In Fig. 5 we plot the weekend-weekday ratio for the individual users. While the overall amount of microblogging activities per day on Sina Weibo is higher on the weekends than during the day, we also discover that 1.2% of the Sina Weibo users perform microblogging activities solely during the weekend (ratio of weekend posts is infinite). For about 50% of the users on Sina Weibo the weekend-weekday ratio is greater than 1 which means that they publish more frequently during the weekend. In contrast, on Twitter we identify only 28% of the users who publish more tweets per day on a weekend than during a weekday.

As depicted in Table 6, the occurrence of organizations and persons is more likely during the weekend than during the week on Sina Weibo whereas locations appear more likely during a weekday. On Twitter, the opposite characteristics can be observed. For example, Twitter users mention locations more frequently during the weekend than during the week. These differences in mentioning entities during weekends/weekdays on Sina Weibo and Twitter respectively may relate to different life styles that Chinese and Western people follow. Investigating the particular reasons for them can be interesting for future work.

Furthermore, we study how individual user interests change over time by calculating the standard deviation of the timestamps of microposts that mention a certain topic (entity). The higher the standard deviation of a certain topic the longer the time period over which the topic is mentioned in the posts. In Fig.6 we plot for each user the average standard deviation of the topics which a user mentioned at least once, and group the average standard deviations by the type of the topics. Overall, we observe that topics on Sina Weibo seem to fluctuate stronger than on Twitter. Sina Weibo users often mention certain concepts only once. For example, for more than 80% of the Sina Weibo users of our sample, the standard deviation of the organization-related topics is 0. These users mention thus organizations only once in their posts. On both platforms the location-related concepts are, on average, mentioned over a longer period of time than organization-related and person-related concepts.

Findings The main findings from the analysis of the temporal behavior (research questions *RQ9* and *RQ10*) can be summarized as follows:

– *F9:* On both platforms, the users posting behavior during weekdays differs the one during weekend: while users on Sina Weibo are more active on the weekends, Twitter users tend to be more active during weekdays. Moreover, user interests change between weekends and weekdays. Again, this change of interests differs between Sina Weibo and Twitter users: while for Sina



(a) Weibo (b) Twitter

Fig. 6. Comparison of topic drift

	China	US
Power distance	80	40
Individualism	20	91
Masculinity	66	62
Uncertainty avoidance	40	46
Long term orientation	118	29

Table 7. Hofstede's cultural index for China and United States

Weibo users we observe a rising interest in persons and organizations during the weekend, the interests of Twitter users focus more on locations. These findings imply that it is beneficial to adapt user interest profiling to the temporal as well as to the cultural context.

– *F10:* User interests change over time. On Sina Weibo, the user interests seem to have a shorter lifespan than on Twitter. Especially, the individual users interests regarding organization-related topics vanish quickly on Sina Weibo while locations feature the longest span of interests.

5 Discussion

Some of our findings can be explained also by cultural differences between the Chinese Sina Weibo users and the Twitter users who are mainly located in the U.S. (more than 80% of the Twitter sample users are located in the United States). According to Hofstede's cultural index [9], people in China can, for example, be characterized by a higher *power distance* than people from the U.S. (see Table 7). This difference might explain our finding *F1* regarding the access behavior (see Sec. 4.1): Sina Weibo users more frequently generate microposts as a byproduct of their other Social Web activities. Therefore, it seems that they are, in comparison to the people who use Twitter, less afraid of disclosing information about themselves. Given the high power distance that is specific to the Chinese culture, we assume that this behavior can be observed because Chinese users do not attribute much impact to their individual activities, i.e. the impact of disclosing information is less because of the high power distance. The more intensive usage of hashtags and URLs which is characteristic for the Twitter users (*F3*, see Sec. 4.2), may relate to both the lower *power distance* and the higher degree of *individualism* of American people (see Table 7). By mentioning a hashtag, microbloggers ensure that their message will appear in the public discussions. Twitter users seem to be more eager to let their posts appear in the public discussion. Hence, they seem to have a stronger belief that their post makes a difference (power distance) and possibly also a higher demand to profile themselves in the public discussions (individualism).

We also observed that Sina Weibo users less frequently mention organizations in their posts than Twitter users (*F5*, see Sec. 4.3). This observation is in line with Hofstede's observation that "employee commitment to an organization is low" in China[6], which is one of the typical indicators for a high *long term orientation*. The sentiment analysis (see Sec. 4.4), which showed that the Chinese

[6] `http://geert-hofstede.com/china.html`

Sina Weibo users are more positive than the Twitter users from the U.S. (*F7*), further supports this cultural difference regarding the long term orientation. In the context of the sentiment analysis, we furthermore discovered that Sina Weibo users are more positively talking about persons than Twitter users (*F8*) which again supports the Chinese tendency for *collectivism* rather than *individualism*.

The temporal analysis (see Sec. 4.5) revealed that Sina Weibo users are less actively publishing microblog posts during the working days and particularly mention less frequently organizations than during the weekend. This can be interpreted as an indicator for *long term orientation* as it implies a rather low commitment for the organization that the user is working for. Sina Weibo users also seem to change their interests rather quickly in comparison to Twitter users (*F10*). While this seems to contradict to the long term orientation of Chinese people, it also reveals that Chinese people adapt faster to new topics which may be interpreted as "an ability to adapt traditions to changed conditions", one of the characteristics of cultures with high long term orientation.

We have given an innovative basis for analyzing microblogging behavior on Sina Weibo and Twitter. Further interpretation and validation of our first set of conclusions can be done in future work, with research questions that follow our conclusions. Independent from these interpretations, we believe that our findings already provide valuable insights for the application of user modeling techniques that are provided by our user modeling framework.

References

1. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: WWW '10, ACM (2010) 851–860
2. Long, R., Wang, H., Chen, Y., Jin, O., Yu, Y.: Towards effective event detection, tracking and summarization on microblog data. In: WAIM'11, Springer (2011)
3. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: CHI '10, ACM (2010) 1185–1194
4. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing User Modeling on Twitter for Personalized News Recommendations. In: UMAP '11, Springer (2011) 1–12
5. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW '10, ACM (2010) 591–600
6. Romero, D.M., Meeder, B., Kleinberg, J.: Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In: WWW '11, ACM (2011)
7. Gao, Q., Abel, F., Houben, G.J.: GeniUS: Generic User Modeling Library for the Social Semantic Web. In: JIST '11, Springer (2011)
8. Mandl, T.: Comparing Chinese and German Blogs. In: HT '09, ACM (2009)
9. Hofstede, G., Hofstede, G.J.: Cultures and Organizations: Software of the Mind, McGraw-Hill (2005)
10. Chen, L., Tsoi, H.K.: Analysis of user tags in social music sites: Implications for cultural differences. In: CSCW '11, ACM (2011)
11. Yu, L., Asur, S., Huberman, B.A.: What trends in chinese social media. CoRR **abs/1107.3522** (2011)
12. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Technical report, Stanford University (2009)